

## Streszczenie

Celem niniejszej rozprawy doktorskiej było opracowanie i optymalizacja metodyki dla obrazowania spektroskopowego w podczerwieni (IR), pod kątem osiągnięcia maksymalnej jakości danych oraz tworzenia modeli klasyfikacji histopatologicznej raka trzustki. Ponadto, podmiotem prowadzonych prac było rozwinięcie technik wyznaczania orientacji makromolekuł, co związane jest z hipotezą mówiącą o dostarczaniu przez nie dodatkowych informacji o chorobie oraz usprawnianiu modeli klasyfikacyjnych. Oba zagadnienia uwzględniają wykorzystanie spektroskopii w podczerwieni, zapewniającej ogrom informacji na temat składu biochemicznego badanej próbki. Co kluczowe, realizacja opisanych powyżej aspektów jest wsparta przez dogłębnie zoptymalizowane procedury wstępnego przetwarzania danych, ze skupieniem na metodach odsumiania oraz korekty linii bazowej.

W związku z nakreślonymi powyżej celami, w pracy pt. „*Obrazowanie w podczerwieni w klasyfikacji histopatologicznej i orientacji makromolekuł, wsparte zoptymalizowanymi metodami wstępnego przetwarzania danych*” można wyodrębnić cztery obszary tematyczne złożone ze zbiorów publikacji. Pierwszy związany jest z szeroko pojętymi optymalizacjami technik odsumiania widm oraz obrazów spektroskopowych, pod kątem podniesienia jakości sygnału oraz poprawy czułości klasyfikacji histopatologicznej. Kolejnym obszarem jest sama klasyfikacja histopatologiczna oraz jej skuteczność w diagnostyce nowotworów trzustki w oparciu o obrazowanie IR. Trzeci obszar badań to wyznaczanie orientacji makromolekuł, w płaszczyźnie (2D) oraz przestrzeni (3D), wykorzystujące spolaryzowaną spektroskopię w podczerwieni, a docelowym zastosowaniem obu metod jest określanie orientacji włókien kolagenowych. Finalnym tematem jest implementacja metody korekty linii bazowej typu EMSC (ang. *Extended Multiplicative Signal Correction*) skierowanej do próbek o cylindrycznym kształcie (np. włókien), wywołujących silne rozpraszanie typu Mie.

W części teoretycznej pracy pokryte zostały podstawy fizyczne zagadnień związanych z badaniami opisanymi w publikacjach wchodzących w skład rozprawy. Rozdział rozpoczął od opisu spektroskopii IR, skupiając się na podstawach spektroskopii oscylacyjnej oraz kwestiach związanych z analizą ilościową. Scharakteryzowany został układ do transmisyjnej Spektroskopii Fourierowskiej w podczerwieni (FT-IR) z uwzględnieniem układu optycznego i aspektów uzyskiwania obrazu. Przybliżono też dwie dodatkowe metody spektroskopii IR, które były wykorzystane w badaniach – mikroskopie QCL (ang. *Quantum Cascade Laser*) oraz O-PTIR (ang. *Optical-PhotoThermal Infrared*). W dalszej części uwaga skupiona została na konkretnych metodach wstępnego przetwarzania danych. W grupie technik odsumiania spektralnego i przestrzennego znalazły się proste filtry (średniej, medianowy, itp.), częstotliwościowy filtr dolnoprzepustowy, filtr Savitzky'ego-Golay'a, Falki, Analiza Głównych Składowych (PCA) oraz metoda Najmniejszej Frakcji Szumu (MNF). Następnie poruszono aspekty korekty linii bazowej, powstającej w wyniku efektów rozpraszania dominujących w transmisyjnym trybie pomiarowym. Główna uwaga została skupiona na algorytmach typu EMSC, wykorzystujących zarówno pełną teorię rozpraszania typu Mie, jak i uproszczenie van de Hulsta. Szczegółowo opisany został algorytm EMSC służący do korekty rozpraszania przez próbki o cylindrycznym kształcie. Trzeci podrozdział wstępu teoretycznego opisuje wykorzystanie spolaryzowanej spektroskopii IR do wyznaczania orientacji dipolowych

momentów przejścia. Opisano metodę 2D, dającą informację ograniczoną do kierunku rzutu momentu dipolowego na płaszczyznę, oraz metodę 3D, jednocześnie wykorzystującą dwa momenty dipolowe i zapewniającą informację przestrzenną o ich orientacji. Poprzez analizę kierunków oscylacji atomów, orientacja momentów dipolowych jest przenoszona na orientację łańcuchów w makromolekułach. Rozdział sfinalizowano opisem metody klasyfikacji Las Losowy (ang. Random Forest), wykorzystanej w badaniach nad modelem histopatologicznym nowotworów trzustki.

Kluczowym dla niniejszej rozprawy fragmentem jest rozdział przedruków. Grupuje on publikacje, będące podstawą pracy, w cztery wspomniane wyżej wątki tematyczne. Każdy z tematów posiada krótki opis kontekstu oraz motywacji do przeprowadzonych badań, jak również kluczowe rezultaty i płynące z nich konsekwencje. Pierwszy i największy zbiór publikacji dotyczył wpływu odsumiania na dane spektroskopowe oraz na skuteczność klasyfikacji histopatologicznej. Dogłębnie zoptymalizowano szereg metod spektralnego i przestrzennego odsumiania w szerokim zakresie ich parametrów. W przypadku zastosowania skierowanego na poprawę jakości sygnału zastosowano symulowane zestawy danych o różnym poziomie szumu oraz rozdzielczości przestrzennej. Umożliwiło to oszacowanie skuteczności metod, zarówno pod kątem poprawy stosunku sygnału do szumu (SNR), jak i zminimalizowania zaburzeń sygnału wprowadzanych przez odsumianie. Uzyskane wyniki wskazały na ogromną skuteczność metod MNF i PCA, uwidaczniając możliwość redukcji czasu pomiarowego do minimum (np. do 4 skanów). Ponadto, skrypty do generacji danych symulowanych, wraz z samymi danymi, zostały udostępnione na zasadach otwartego dostępu. Odszumianie zostało również zoptymalizowane pod kątem poprawienia skuteczności klasyfikacji z wykorzystaniem Lasów Losowych. Pokazano, że po zastosowaniu metody PCA lub MNF do danych FT-IR, czułość klasyfikacji na poziomie piksela wzrasta o nawet 43-44%. W przypadku klasyfikacji opartej o dane pochodzące z mikroskopii QCL, zastosowanie MNF wskazało możliwość redukcji informacji spektralnej do zaledwie 20 zmiennych (pasm). Przytoczone optymalizacje odsumiania miały kluczowy wpływ na opisaną dalej klasyfikację.

W ramach tematyki związanej z klasyfikacją histopatologiczną trzustki, celem badań było stworzenie modeli umożliwiających detekcję nowotworów w oparciu o spektroskopię FT-IR oraz metody uczenia maszynowego (Lasu Losowego). W tym celu zebrano dane dla ponad 600 biopsji trzustki, pochodzących od około 250 pacjentów, wykorzystując zarówno standardowy (15x) jak i wysokorozdzielczy obiektyw (36x), zapewniający bardziej szczegółową informację przestrzenną. Finalnie stworzone zostały modele wyróżniające po 6 klas, co umożliwia detekcję nowotworu lub zapalenia ze skutecznością na poziomie 95%. Dodatkowo, zaprezentowano możliwość detekcji marginesu chirurgicznego uwidaczniającego wyraźną granicę pomiędzy zdrową, a nowotworową tkanką.

Teorie na temat wpływu tkanki włóknistej, będącej mikrootoczeniem nowotworów, na przebieg choroby oraz skuteczności leczenia były motywacją do rozwoju metod wyznaczania orientacji makromolekuł, w oparciu o spolaryzowaną spektroskopię FT-IR. Przedstawiono rezultaty zastosowania metody 2D do tkanek bogatych we włókna kolagenowe, a otrzymane wyniki wyraźnie wskazały kierunki orientacji molekuł. Co więcej, funkcja Hermansa wyeksponowała wysoko i nisko ustrukturyzowane obszary, odpowiednio dla tkanki włóknistej oraz komórek nabłonkowych. Było to pierwsze zastosowanie tej metody do tak skomplikowanego układu jak tkanka. W dalszym kroku zaprezentowano pierwszą

implementację do danych eksperymentalnych, teoretycznej metody wyznaczania orientacji 3D. Wykorzystano w tym celu dobrze opisany w literaturze układ polimeru polikaprolaktonu w krystalicznej formie sferulitu. Metodę zaaplikowano również do danych O-PTIR, umożliwiających pomiary z submikronową rozdzielczością. Wyznaczona orientacja makromolekuł była zgodna z przewidywaniami teoretycznymi, tworząc współśrodkowe okręgi wokół centrum nukleacji, co przybliżyło potencjał tej metody w nietomograficznym i niedestruktywnym wyznaczaniu orientacji 3D.

Tkanka włóknista analizowana w ramach badań nad orientacją ujawniła wysokie tendencje do rozpraszania światła, w znacznym stopniu zaburzające analizy. Dlatego czwartą i ostatnią poruszoną w rozprawie tematyką jest korekta rozpraszania. Przedstawiono implementację algorytmu EMSC służącego do korekty rozpraszania typu Mie przez próbki o cylindrycznym kształcie. Algorytm rozszerzono również o przypadek liniowej polaryzacji światła, aby jak najlepiej dopasować go do zastosowań związanych z wyznaczaniem orientacji. W celu przyspieszenia obliczeń wykorzystano procesory graficzne, a stworzona w oprogramowaniu MATLAB biblioteka została udostępniona na zasadach otwartego dostępu. Włókno polimeru o silnych właściwościach rozpraszających zostało wykorzystane do walidacji nowego algorytmu, który okazał się skuteczniejszy od metody referencyjnej.